

Spitzer Heritage Archive

Xiuqin Wu, Trey Roby, and Loi Ly

*IRSA/SSC*¹, *California Institute of Technology, 100-22, Pasadena, CA, USA*

Abstract. The Spitzer Heritage Archive will host all the raw and final re-processed science and calibration data products from the observations made by Spitzer Space Telescope. The interactive web interface will give users the tools to search the database and explore their search results interactively. We also reuse the existing software and services and pay close attention to the re-usability of the newly developed system, making it easy to expand and adopt new technology in the future. This paper discusses our design principles, system architecture, reuse of the existing software, and reusable components of the system.

1 Introduction

Spitzer Space Telescope² is the fourth and final element in NASA's family of Great Observatories. Spitzer was launched in August 2003. It observed in the 3.6 to 160 micron range with three instruments: IRAC, IRS, and MIPS for five and half years. The cryogen was depleted in May 2009 and it started extended warm mission in September 2009, observing in 3.6 and 4.5 microns using IRAC. The Spitzer observations have produced vast amount of imaging and spectroscopy data. The heritage archive will make this treasure trove available for the astronomy community. IRSA³ is NASA's InfraRed Science Archive operated in IPAC, Caltech. It currently hosts all the IRAS and 2MASS data, and will be the final home of Spitzer Heritage Archive. The official handover will be April 2011.

To satisfy the requirements and meet the expectations of our users, we looked into the new technologies in the web interface development. We believe that a good user interface will enhance the user experience to explore the Spitzer data, thus increasing the potential of science discovery. We decided on using Google Web Toolkit (GWT) to deliver an AJAX based web interface, which is very powerful and easy-to-use. It will give users the tools to search the database and explore their search results interactively. The meta data will be presented in an easy to read table format. Users can page through their data results, perform sorting, and filter on any column. They can also view the FITS images and spectrum plots before they decide to download them. We also reuse the existing software and services, pay close attention to the re-usability of the newly developed system, make it easy to expand and adopt new technologies in the future. In the following sections, we discuss our design principles, system

¹ Spitzer Science Center

architecture, reuse of the existing software, and reusable components of the system.

2 Spitzer Data Products

Spitzer Heritage Archive will host about 25 TB of data when it is handed over to IRSA. The data products include:

1. Spitzer raw data
2. Spitzer pipeline processed data:
 - Science quality data products
 - Ancillary data
3. Calibration products
4. Data quality analysis information
5. Observation design information
 - Program title & abstract
 - Observational constraints
6. Legacy data
 - Including images, spectrum, catalogs, models, etc.
7. SSC-produced “enhanced” data
 - Super mosaics generated by combing IRAC and MIPS data together
 - Source list extracted from the super mosaics

Spitzer Heritage Archive will also eventually contain the IRAC data from the “warm” mission, which will be added incrementally.

3 Design Principles

Spitzer Heritage Archive will be accessible through the web. We want to deliver a well designed and easy to maintain software system. We paid special attention to the following design principles to make this system user friendly:

1. It is easy to learn and use.
2. It responds quickly.
3. It is well integrated.
 - Access to other online images and catalogs
4. It has a consistent look and feel.
 - Same way to display error message
 - Same layout for buttons
 - Same color for the same type of dialog
 - Same placement for help button

We strive to develop a system with a strong object oriented design, which makes it easy to maintain, add new functionalities, adapt to new technologies, and be reused for other systems.

² <http://spitzer.caltech.edu>

³ <http://irsa.ipac.caltech.edu>

4 Major Features

We derived the major features of our system based on the feedbacks from Leopard users and the science usecases from Spitzer and non-Spitzer users.

4.1 Search Capabilities

The interface will provide the basic and Spitzer specific search functions to explore the Spitzer Heritage Archive:

- Radius search with instrument and wavelength constraints, for single position or a list of positions
- Observer (PI) search
- Program ID search
- Observation Request (AOR) ID search
- NAIF ID (for moving target) search
- Campaign ID search
- Observation date search
- Free form text search (Google like) – user can search any text in proposal abstract, proposal title, and category

4.2 Data Presentation

To efficiently use the limited screen space, we decided to present the different levels of data products in different tabs. Users can choose what levels of data they desire for each search and they can see all of their choices by tabbing through them.

To provide the user with more information about the search results, we chose to display the data in two tabbed panels side by side - the left panel displays meta data in an easy to read tabular form, and the right panel displays more information related to the selected row of meta data according to the tabs selected. Those two panels work together to let the user explore the results interactively. Currently there are three tabs for more information: detailed meta data not displayed in the table, an image or spectrum plot, and the outlines of all the displayed data on a user selected image background.

4.3 Interactive Data Exploration

With the AJAX technology, we can provide an interface to let users manipulate and explore the data more interactively, allowing them to know the data more before making a decision on downloading. See Figure 1. for a screen shot of the tabular meta data, the selected image, and the overlay of the displayed data coverage area.

- Table manipulation - Users can page through the data, sort on any column, filter on any column, customize which columns to be displayed, save the data.
- Image Display - If the selected row of meta data is associated with a FITS image, the image will be displayed. Users can get the position and flux value readout by moving cursor over the image, change the color and stretch of the image, perform zoom in and out, put coordinate grid over the image. Users can also rotate the image, display north and east arrows on the image, and get a close-up view using the magnifier.

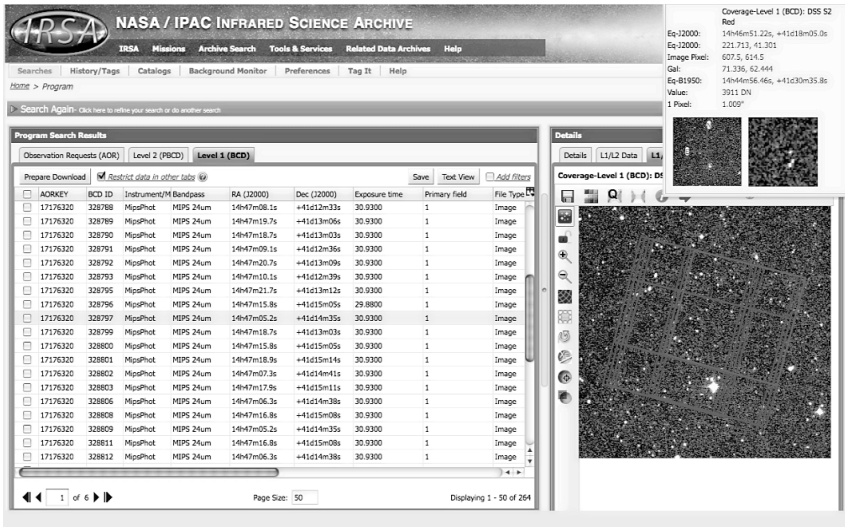


Figure 1. Screen Shot of Spitzer Heritage Archive.

- **Spectrum Display** - If the selected row of the meta data is associated with spectrum data, the flux and wavelength will be graphed. Users can get the flux and wavelength values readout by moving cursor over the graph. User can also zoom in on the graph in the future.

4.4 Data Download

Our goal is to allow users to download any data anytime. While exploring the data interactively, users can select one or more data products to download by clicking the check box for each data row. Users can also download all the data by clicking the check box at the header. For individual data the download is instant. For a large amount of data, the system will package them into one or more gzipped files. While the system packages the selected data for download in the background, users can continue to do more searches, more data exploration, and more data download. Users can monitor the data packaging progress status using the background job monitor. Users can also choose to ask the system to send an email reminder when all the data is packaged and ready to download.

Users can download different levels of data at different meta data tabs. Users have choices to download the main data products along with the uncertainty and mask files, the ancillary data products associated with the main data products, and the calibration data files.

4.5 User Registration

We provide optional user registration. The system will remember the user's email for download notification, the search history, and the user preferences for registered users between different sessions. For non-registered users, some of

those can be accomplished by cookies. However, if users change computer or clean out the cookies, those will be lost.

In the future, user registration will be required to access the restricted data, like the proprietary data from Spitzer “warm” mission.

4.6 Two Ways to Access

We will provide web interface and application program interface to access the Spitzer Heritage Archive. The web interface will provide users with an interactive way to search, explore, and download data. The application program interface will enable users to write their own programs and interfaces to search the database and download the data.

4.7 Search History and Tagging

Users can see their search histories. They can mark their favorite searches, modify the description of the search, and resubmit the search. The tagging function will allow users to tag the search and come back to the same data result page by using the tag name or URL provided.

5 System Architecture and Reusability

5.1 Major Technology

- Multi-tiered architecture to provide the security, scalability, and stability. The Apache web server is what users know about. Its main job is to get the user requests to the proper Tomcat application server behind another firewall. The application servers are configured the same and can handle all the tasks. Any one of them can be replaced without much impact on the whole system. We can also scale up by adding more application server to deal with increased demand on the system. The file servers and database servers will host the data files and provide DBMS services. Any of them can be upgraded without impact on other parts of the system. See Figure 2. for system deployment diagram.
- AJAX technology to provide the interactive web interface. We chose Google Web Toolkit (GWT) to develop the interactive user interface. It provides us with a very nice framework to use. Please see Trey Roby’s paper (2010) for more information on this topic.
- generic data manipulation to provide the three layered design. Each layer has its clear defined responsibilities. By encapsulating the code in each layer, we can isolate changes and minimize the impact one layer has on others. This will allow us to have a more secure, scalable, and stable system. The persistence layer provides the uniform access to external data in files, DBMS, etc., making it easy to adapt to new database technologies in the future. The logic layer provides the generic data manipulation, isolating the astronomy related data manipulation here. The presentation layer receives user requests and generates results, making it easy to change the data display format. See Figure 3. for system architecture diagram.

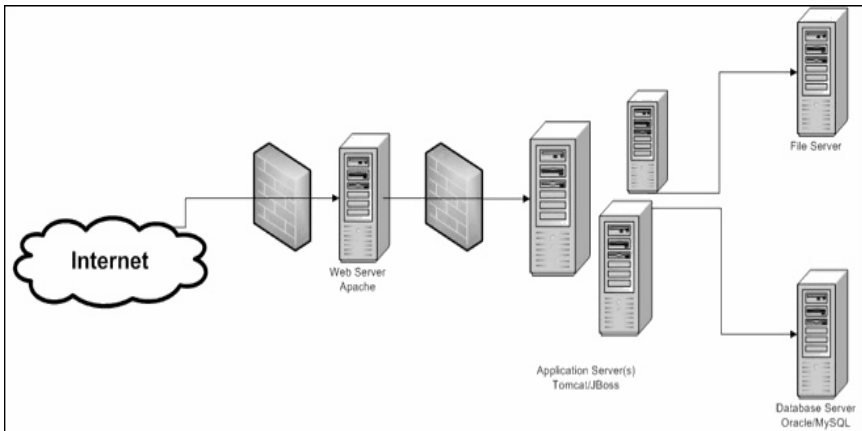


Figure 2. System Deployment Diagram.

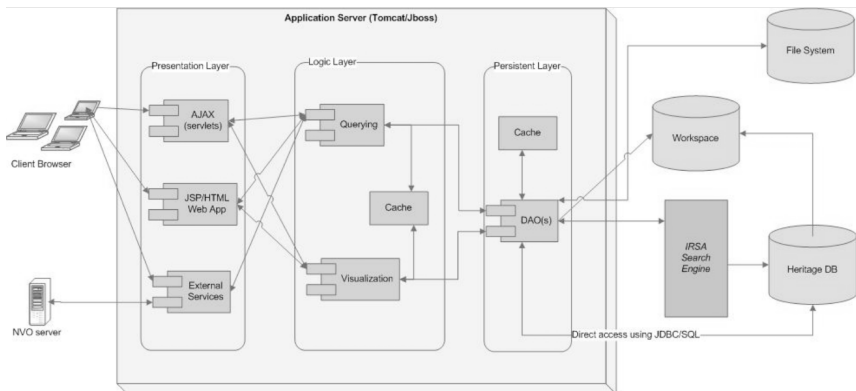


Figure 3. System Architecture Diagram.

5.2 Performance Enhancement

Performance is very important for a good user interface. We explore all possible ways to enhance the performance of the system.

- Maximize use of the client computation power. Today's laptops and desktops are very powerful hardware. By using AJAX technology, we can transfer many tasks to run on the client machine, like input validation, name resolution through NED or Simbad, coordinate conversion, and FITS image position readout.
- Fast image search using R-tree indexing technology.
- Cache as much as possible.
- Partial return of large data results. User can start examining the data before all the results return.

5.3 Existing Software Reuse

We have developed a large amount of Java classes for Spitzer planning and proposal submission tool (Spot) and the current application to access Spitzer archive (Leopard) in the last ten years. Choosing GWT technology enable us to reuse those classes, which have been tested and validated by SSC integration and test team, more importantly by hundreds of users in more than seven years and six proposal call cycles. Some of the classes name resolution through NED or Simbad, coordinate conversion, FITS image display, and getting images and catalogs from other archive. We are also able to reuse the mature catalog and image searching software developed in IRSA in last ten years.

5.4 Future Software Reuse

We want to develop a system that can be easily adapted for other archive systems. At the very beginning we designed a common web frame work that is not Spitzer specific. We named it Firefly. The Spitzer Heritage Archive is built upon Firefly, which includes the infrastructure and many common building blocks:

1. Tabular data display and table manipulation
2. FITS image display
3. Spectrum plot
4. Data download management
5. Input fields and data validation: RA, Dec, radius, etc.
6. Coordinate conversion

6 Schedule

We plan to have a public beta release of the Spitzer Heritage Archive system in February 2010, the first public release in July 2010, and final handover to IRSA for the management of the system in January 2011.

References

Trey Roby 2010, in ASP Conf. Ser. 434, ADASS XIX, ed. Y. Mizumoto, K.-I. Morita, & M. Ohishi (San Francisco: ASP)